Coordination Strategies of Chimpanzees and Human Children

in a Stag Hunt Game

Supplementary Methods and Results

Subjects:

Experiment 1 Low Risk Game with children

Pilot testing with children aged two to four years indicated that 4 years was the youngest age that they were able to reliably complete the training and pre-test requirements (2.5 year olds *N*=12; 3.5 year olds *N*=10).

Experiment 2a High Risk Game with chimpanzees

Table S1: Subject information for Exp 2a (high risk game with chimpanzees), including their participation in the low risk game (Bullinger et al. 2011).

Name	Age (years)*	Sex	High Risk Game Group (Exp 2a)	Low Risk Game Group
Lome	10	Male	No-barrier first	Barrier-first
Robert	35	Male	No-barrier first	Barrier-first
Natascha	32	Female	No-barrier first	No-barrier first
Pia	12	Female	No-barrier first	No-barrier first
Tai	10	Female	No-barrier first	NA
Frodo	17	Male	Barrier-first	NA
Lobo	7	Male	Barrier-first	No-barrier first
Fraukje	35	Female	Barrier-first	No-barrier first
Sandra	18	Female	Barrier-first	Barrier-first
Ulla	32	Female	Barrier-first	Barrier-first
Dorien	31	Female	NA	NA
Riet	33	Female	NA	NA
Swela	18	Female	NA	NA

NA: subjects did not pass training phase and therefore did not take part in the test trials
**at time of Exp. 2a*

Experiment 2b High Risk Game with children

The total number of subjects was increased from 40 in Experiment 1a to 48 in Experiment 2b because the number of pairs was counterbalanced for an extra factor, pre-test order, in Experiment 2b, as well as sex and condition order

Procedure

The following section provides further details of the procedures for the training and pre-tests phases in each of the studies. However, the children also had to pass the same pre-tests to show that they understood the pay-off structure.

Experiment 1: Low risk game with 4 year olds

The amount of training and testing trials was reduced for the children so that training and testing could be completed on a single day.

Warm-up. Experimenter 1 (E1) and Experimenter 2 (E2) introduce themselves to the children and play together.

Children were trained separately by E1 across a series of trials so that they understood the apparatuses.

Hare training. Frist they were trained on their own hare box in two trials. They learned to collect the low value food (rice puffs) from the box, and also that once closed it could not be opened again.

Stag training. Children learned to operate the stag box with E1 across three trials, so that they understood that two people were required to operate it; that the stag rewards were available for a short time only; and that sounds indicated the presence of the stag.

Game training. Intermixed with more hare-only trials were trials in which both the hare and stag were available. This included situations in which their partner was either: available, available but delayed, present but unavailable (busy doing another task) or not present. The game training also included a conversation part, in which E1 asked the children a series of question to highlight key features of the game (i.e. what they can collect individually, and what they can collect cooperatively).

Pre-tests. Once children had completed training trials, they were presented with pre-tests to check whether they had understood the pay-off structure of the game from the training trials. Pre-tests consisted of two types of trials: *social pre-test* and *non-social pre-test* (details below). To pass the pre-tests participants were required to respond correctly in two out of three trials for both the social and non-social trial types. The social pre-test trials also function to ensure that they value the stag over the hare. All subjects completed this pre-test successfully prior to each testing day, so we can be confident that at the time of testing the stag was of higher value.

Social pre-test. E1 and one child entered the test room, in which the stag box and the child's hare box were set-up. While the child began to collect the hare, E1 waited at the stag box and the stag rewards appeared. The child then had 15 seconds to decide whether or not to stay at the hare. They were considered to have passed if they leave the hare and pull the rope at the stag with E1 within 15 seconds.

Non-social pre-test. E1 and the child entered test room, set up as in the social pre-test with the stag box and the child's hare box. When the child had begun to collect the hare, E1 made an excuse and left the room. At this point the stag appeared for 15 seconds, after which it was retracted. Children were considered to have passed the non-social pre-test if they stayed at hare and did not attempt to retrieve the stag.

Experiment 2a: High risk game with chimpanzees

Training: stag. The aim of the stag training was to ensure that subjects a) understood a partner was required to access the stag reward and b) could wait for a partner before pulling the rope alone (and thus causing the reward to be inaccessible). For each training trial the stag reward was placed on the apparatus and the subject entered the test cage. A partner entered the adjacent cage after a delay of increasing intervals (5, 10, 20, 30 seconds). To complete this training phase subjects were required to acquire the reward on two consecutive trials at each time interval, within the same testing session (this phases could be passed in a minimum of eight trials). If they did not acquire the reward on three consecutive trials they received one motivation trail (partner present without delay) before beginning again with 5 second delay trials. All subjects completed this phase with a human experimenter and then again with a conspecific stooge partner. Subjects required an average of 17.6 trials (range: 8-81 trials) to pass with a human partner and 12.6 trials (range: 8-42 trials) with a conspecific partner.

Training: hare. Only the subject's hare was available. To be successful in this phase subjects had to a) stay at the hare box (without letting the door close) for at least 60 seconds (or, until the juice was finished) and b) experience the irreversible closing mechanism of the door by trying to re-open it on at least one occasion. Subjects experienced a) and b) in both testing cages. Subjects completed this phase in an average of 14 trials (range: 3-43).

Training: partner but no stag. Participants entered the testing room to see the ropes of the stag are extended into the testing rooms and a stooge (with no hare) in the adjacent room. However, there is no reward available at the stag apparatus. To succeed on this trial the participants had to go to their hare without pulling at the rope and stay there for 60 seconds (or until the hare box was empty) on three consecutive trials, this was repeated in both testing cages. The average number required to pass in both cages was 12 trials (range: 6-23).

Training: partner with hare. Subjects did not have access to their hare, or the stag but observed a stooge in the adjacent cage retrieving their hare reward (for at least 60s, or until the hare was empty) on 3 trials for each testing cage.

Pre-test.

Social pre-test. In the social pre-test the stag apparatus, as well as the hare box in the subject's cage was set-up. The subject and the stooge partner each entered their cages and, when the subject had started to drink their hare reward, the stag reward (6cm of banana each) appeared. The stag was available for 20s until a rope was pulled from outside the testing room by E2 and the banana fell to the floor and out of reach. As the auditory cues indicating the appearance of the stag in Bullinger et al. were removed in the current procedure, subjects' names were called in earlier trials and sessions to encourage them to check the stag. To pass the social pre-test subjects were required to leave the hare and go for the stag within 20s in 4 out of 5 trials (only sessions without any auditory cues). Subjects passed the social pre-test in an average of 8.09 sessions (40.50 trials, range: 2-20 sessions, 10-100 trials).

Non-social pre-test. The set-up was the same as the social-pre-test, except that there was no partner present in the adjacent cage. To pass the non-social pre-test subjects had to stay at the hare in 4 out of 5 trials. Subjects passed the non-social pre-test in an average of 2.78 sessions (13.89 trials, range: 1-6 sessions, 5-30 trials).

When subjects had reached criterion for the social and the non-social pre-test they were presented with blocks of 5 trials of both types on the same day and were required to pass 4 out of 5 trials in both blocks. Subjects passed the combined pre-test in an average of 5.18 sessions (84.36 trials, range: 3-11 sessions, 30-213 trials).

Test.

Refresher pre-test. Prior to testing, subjects were required to pass 'refresher' pre-tests on the day of testing. These consisted of pairs of one social and one non-social pre-test trial. If subjects responded correctly (according to the criteria described above) in either the first pair, or the second pair, they could go on to test trials that day. If they did not, they took part in a third pair of pre-test trials as training and were given the pre-tests again the following day. If their partner did not pass the refresher pre-tests that day, both were given the pre-tests again in the following session. Subjects required an average of 1.65 sessions to pass the pre-test (range: 1-13). As partners did not always pass the pre-test on the same testing day, the average number of pre-tests sessions between testing days was 2.56 sessions.

Test trials. In a test trial both hares and the stag apparatus were set-up. The subjects entered their respective cages and, once they had both opened the doors to their hare boxes, the stag rewards were placed on the platform by E1and remained there for 20s. If the subjects had not retrieved the stag at 20s (by simultaneously pulling on the rope), E2 pulled a second rope attached to the stag platform from outside the testing room (unseen by the subjects), thus causing the banana pieces to fall on the floor out of reach. Subjects received 6 test trials per testing day. If either partner did not participate (drink form hare) on 3 consecutive trials, the session was ended and completed in the following testing session.

Experiment 2b: High risk game with 4 year old children

Warm-up. Experimenter 1 (E1) and Experimenter 2 (E2) introduce themselves to the children and play together with toys in the warm-up room.

Demonstration: Unsuccessful attempt at Stag. In this trial children see that the stag cannot be acquired alone, and that the rewards disappear after some time. Both children joined E1 and E2 in the test room, with only the stag box set-up (E3 sits behind the box pretending to read, and did not interact with the children). The stag rewards appeared (two

opaque red plastic balls) and the children watched E1 fail to obtain them alone by pulling the rope on her side of the box, while E2 was occupied with another task. After the rewards disappear again E1 tells E2 what happened, then the children leave the room with the experimenters.

Training: Successful attempt at Stag. Children took part in a successful attempt to acquire the stag with E1. Each child entered the test room with E1 individually (so that they did not have any experience cooperating with their partner) to find the stag box set-up. E1 and the child played with a ball while looking out for the stag reward. Once it arrived, they stopped playing and had to figure out how to get the reward: first E1 and the child pulled their ropes individually. When this did not work they tried pulling simultaneously and they were able to reach the reward and open the balls to find out that there was a gummibear inside. E1 and the child leave the room to store their reward in their collection boxes.

Training: hare. The aim of this trial was that the children understood the mechanism of the hare door and that they practiced extracting the hare reward. Again, both children entered with E1 and E2, both hare boxes were set-up and the children sat in front of what would be their hare box (indicated by a coloured cushion). They observe E1 and E2 collecting the hare, before collecting some themselves. Both experimenters dropped the doors of the hare boxes and the children were encouraged to attempt to re-open the box, thus experiencing that they could not. The children then left the room with the experimenters to store the rewards they had collected.

Pre-tests. Social and non-social pre-tests were as in Exp. 1a, but the structure of the pre-test section differed slightly. Children were first presented with blocks of social and non-social practice pre-tests. The number of practice trials was determined by when they responded correctly (with a maximum of four trials in each block). In the non-social trials

they were considered to have passed if they left the hare and pulled the rope at the stag with E1 within 15 seconds. Children were considered to have passed the non-social pre-test if they stayed at hare and did not attempt to retrieve the stag, or , if they left hare to go and get E1 from outside the room. In this case E1 would be too slow to react to the request so that children never successfully acquired the stag in non-social trials.

After practice trials, children were presented with up to two pairs of one social and one non-social pre-test trial. If they responded correctly to both trials, in either the first or the second pair, they could go on to test trials with their partner. If either partner failed the pre-tests, neither went on to test trials.

Instructed pre-tests. When it was apparent that children were struggling to pass the pre-test (in 7 out of 16 pairs tested up to this point, at least one child did not pass) instructed pre-tests were introduced as a final set of pre-tests. If they passed this pair they also went on to the test phases. The instructed social pre-test trial was identical to the social pre-test, with the addition of E1 indicating the arrival of the stag and requesting the child to pull the rope with them if they did not immediately join them ("Da sind Gummibärchen. Komm, zieh mit"). The instructed non-social pre-test trial was as the standard non-social except that E1 instructed the child to stay at the hare until she returned ("Blieb bei den Schokopops bis ich zurück komm"). Five children required the instructed pre-tests, three of which passed and were included in the final data set.

Coding

Experiment 2a: High risk game with chimpanzees

Note that the definitions of successful coordination and communication differ slightly from that used in Bullinger et al. Prior to analysis, data from Bullinger et al. was recoded in accordance with the current definitions.

Experiment 2b: High risk game with children

Type of communication. The coding scheme of the content category also included the categories: *questions* to their partner relating to objects or actions in the game (e.g. "are you coming?"); and simple one-word *answers* to these questions (e.g. "yes"). Each occurrence of communication was coded for the type of communication hierarchically. If an individual used content communication and attention-getters in the same phase it was coded as content. Within the content category the imperatives were prioritised over informatives, and informatives over questions and answers. There were very few occurrences of questions and answers only (Exp. 1a: no questions/answers; Exp. 2b: questions 0% communication at hare, 3.45% at stag and answers 2.86% at hare and 3.45% at stag), all further analysis of content type excluded these cases and focussed on the use of imperatives and informatives.

Analysis

We used Generalised Liner Mixed Models (GLMMs; [1]) with binomial error structure and logit link function for all analyses. Analyses were conducted in R version 3.0.2 [2] using the function glmer of the lme4 package [3].

Prior to inspection of the model results, we conducted an overall test of the full model (with all test and control variables) compared to the null model (the full model without test variables) using a likelihood ratio test [4]. Only when this was significant did we consider the model results. The significance of test variables were tested individually using drop1 function of the lme4 package (using a likelihood ratio test). If the model included an interaction that was not found to significantly contribute to the model it was removed to produce a reduced model.

The models all included the random effects of subject and pair (in the case of analyses on a trial basis only pairs were included). In order to reduce model complexity we tested the

contribution of several random slopes (for the analysis of individual decisions this was subjects across conditions, subjects across trials, pairs across conditions and pairs across trials; and for analysis on a trial basis this was: pairs across conditions and pairs across trials) and only those that contributed to the model were included. For measures of behaviour at stag (communication, monitoring) only trials in which individuals left hare were included in analysis.

Below are details of each of the models relating to results reported in the respective results sections of the main article. The p-values reported are those of the Chi-square tests of the individual terms.

Models 1-6 analyse the data from Exp. 1 (low risk game with children); models 7-11 analyse the combined data from Bullinger et al. and Exp. 2a (low and high risk games with chimpanzees); finally, models 12-20 analyse the combined data from Exp. 1 and 2b (low and high risk games with children). We do not statistically compare as differences between species in the relative values of hare and stag and experimental design (e.g. pairings, trials and session number) would make the results difficult to interpret in a meaningful manner.

Experiment 1: Low risk game with 4 year old children

Model 1: Decisions to leave hare. This model examined the likelihood to decide to leave the hare. The dependent measure was whether or not an individual left their hare on a given trial (six trials per individual). We were interested in whether the presence of the barrier affected children's decision making, and expected that this could be effected by the order of conditions (i.e. whether they have already coordinated with their partner in the no-barrier condition). Therefor the interaction between condition and order was included in the model. The overall test of the full model (including fixed factors: the interaction between condition and order, sex, trial number; random effects: individual and pair) compared to the reduced

model (the same as the full model excluding test variables: condition, order, sex) indicated that the test variables did not significantly contribute to the likelihood to leave hare (χ^2=3.28, *df*=3 , *p*=0.35, *N*=240).

Model 2: Coordination success. This model investigated the factors affecting coordination success, thus we analysed the coordination success of pairs (6 trials per pair). For the same reasons as with the previous model we included the interaction between condition and order. The overall test of the full model (including fixed factors: the interaction between condition and order, trial number; random effects: pair) compared to the reduced model (the same as the full model excluding test variables: condition, order and sex) indicated that the test variables did not significantly contribute to the likelihood to coordinate successfully (χ^2=2.27, *df*= 3, *p*=0.52, *N*=120).

Model 3: Communication at hare. Model 3 examined the likelihood of children to communicate while still at hare. The dependent variable was whether or not an individual made game relevant verbal communication while still at hare (six data points per individual). The overall test of the full model (including fixed factors: the interaction between condition and order, sex, trial number; random effects: individual and pair) compared to the reduced model (full model excluding test variables: condition, order, sex) suggested that the test variables did not significantly contribute to the likelihood to communicate at hare (χ^2= 5.81, *df*= 7, *p* =0.56, *N*=240).

Model 4: Communication at stag. Model 4 examined the likelihood of children to communicate after leaving hare. The dependent variable was whether or not an individual made game relevant verbal communication after leaving hare (only trials in which an individual left hare were analysed). The overall test of the full model (including fixed factors: the interaction between condition and order, sex, trial number; random effects: individual and pair; and random slopes: subjects across trials) compared to the reduced model (full model

excluding test variables: condition, order, sex) indicated that the test variables did not significantly contribute to the likelihood to communicate after leaving hare (χ^2= 4.26, *df*= 3, *p* =0.23, *N*=234).

Model 5: Monitoring of partner at hare. Model 5 investigated children's use of visual monitoring while still at hare (the partner could be either still at hare, or also at stag). We tested the likelihood to look towards their partner's head, when the child was still at hare across conditions (they could monitor in the barrier condition by looking around the barrier), including the interaction between condition and order of conditions in the model. The overall test of the full model (including fixed factors: the interaction between condition and order, trial number; random effects: subject, pair) compared to the reduced model (the same as the full model excluding test variables: condition, order and sex) indicated that the test variables did not significantly contribute to the likelihood to monitor their partner at hare (χ^2=2.29, *df*= 4, *p*=.68, *N*=240).

Model 6: Monitoring of partner after leaving hare. Model 6 investigated children's likelihood to monitor their partner after leaving hare (in which case the partner could be either still at hare, or also at stag). The overall test of the full model (including fixed factors: the interaction between condition and order, trial number; random effects: subject, pair) compared to the reduced model (the same as the full model excluding test variables: condition, order and sex) indicated that the test variables did not significantly contribute to the likelihood to monitor their partner at hare (χ^2=2.63, *df*= 4, *p*=0.15, *N*=240).

Experiment 2a: High risk game with chimpanzees

Model 7: Decision to leave hare. Model 7 investigated the decisions of individuals to stay at hare or go for stag (24 decisions per subject). We expected the chimpanzees would respond to the barrier condition differently in the high risk game, and potentially also

depending on the order these conditions were experienced in. Thus, we included all possible interactions between game, condition and order in the full model. We also tested for the effect of trial number to examine whether coordination improved over time. Age and sex were initially included in the full model, but were removed when they were found not to significantly contribute to reduce model complexity. The overall test of the full model (including fixed factors: game (high or low risk), condition, order, partner number and trial number; random effects: subject, pair) compared to the reduced model (full model excluding test variables: game, condition, order and trial number) suggested that the test variables did significantly contribute to the likelihood to leave hare (χ^2=177.09, df=8, p<0.001, N=1536). The results indicate a significant three-way interaction between the game, condition and order (Table S2).

Table S2: The full results of the model of decisions to leave hare.

Term (Test category)	Estimate	SE	χ^2	df	p
Intercept	6.32	1.15			
Partner number	0.01	0.25	0.00	1	.96
Trial number	0.11	0.04	8.02	1	<.01
Game (high risk)	-5.99	1.00			
Condition (no-barrier)	2.28	1.57			
Order (no-barrier first)	-5.02	1.16			
Game*Condition	-2.12	1.64			
Game*Order	5.78	1.46			
Condition*Order	-3.37	1.75			
Game*Condition*Order	4.86	2.14	5.79	1	.02

The table presents the effect of test category in relation to the following reference categories: Game-low risk; Condition-barrier; Order-barrier-first. These baseline categories are valid for all subsequent models.

Model 8: Coordination success. Model 8 examined the likelihood of coordination success of pairs of chimpanzees (24 trials per pair). As with the previous analysis, we included all possible interactions between game, condition and order in the model. We did not include control factors for sex or age, because analysis was on the behaviour of the pair and not the individual (and neither age nor sex significantly contributed to individual likelihood to leave hare in model 8). The overall test of the full model (including fixed factors: game (high or low risk), condition, order, partner number and trial number; random effects: pair) compared to the reduced model (the same as the full model excluding test variables: game, condition, order and trial number) suggested that the test variables did significantly contribute to the likelihood to leave hare (χ^2=125.84 , *df*=8, p<0.001, *N*=768). The results mirror those found with model 6: a significant three-way interaction between the game, condition and order (Table S3).

Table S3: The full results of the model coordination success.

Term (Test category)	Estimate	SE	χ^2	*df*	*p*
Intercept	5.77	1.19			
Game (high risk)	-6.28	1.28			
Condition (no-barrier)	2.35	1.80			
Order (no-barrier first)	-5.54	1.47			
Trial number	0.11	0.05	5.84	1	.02
Game*Condition	-2.19	1.88			
Game*Order	5.80	1.90			

Condition*Order	-3.44	1.98			
Game*Condition*Order	5.58	2.32	6.63	1	.01

Model 9: Coordination on hare. While coordinating on hare is the pay-off dominant solution, there is a second coordinate solution, which is for both individuals to stay at hare. It is possible that this is the way chimpanzees were solving the high risk game. To analyse this we recoded coordination success so that coordination failure included only trials in which one individual stayed at hare and the second left for the stag (i.e. coordination success included both coordination at hare and coordination on stag) and analysed coordination failures across games, conditions and order of conditions (including all two and three-way interactions). The overall test of the full model of the likelihood of coordination failure (including fixed factors: game, condition, order, trial number; random effects: pair; random slopes: pairs across conditions and trials) compared to the reduced model (full model excluding test variables: game, condition, order and trial number) suggested that the game type did significantly contribute to the likelihood of a coordination failure (χ^2=55.76, *df*= 8, *p*>0.001, *N*=768). Table S4 summarises the full model.

Term (Test category)	Estimate	SE	χ^2	*df*	*p*
Intercept	-4.05	0.90			
Game (high risk)	3.53	0.94			
Condition (no-barrier)	-0.76	0.27			
Order (no-barrier first)	2.11	0.16			
Trial number	-0.06	0.04	2.07	1	.15
Game*Condition	0.68	0.31			
Game*Order	-2.46	0.44			

Condition*Order	2.05	0.45			
Game*Condition*Order	-3.70	0.55	5.82	1	.02

Model 10: Communication at stag. Model 10 examines the likelihood of communication while leaders wait at stag. Only trials in which at least one individual left hare were included in the analysis. As there were few trials with communication, only the main effects for game, condition and order were included in the model. We also included a term for the time waiting for a partner at hare (the time between leaving hare and the end of the trial) as we expected that with increased waiting time individuals would be more likely to communicate to a partner. The overall test of the full model (including fixed factors: game, condition, order, time waiting at stag, and trial number; random effects: pair; and random slopes: pairs across conditions and trials) compared to the reduced model (the same as the full model excluding test variables: game, condition, time waiting at stag and order) suggested that the test variables did significantly contribute to the likelihood to communicate at stag (χ^2=37.41, *df*=5, p<0.001, *N*=686). Table S5 provides details of the full model.

Table S5: The full results of the model of communication as stag.

Term (Test category)	Estimate	SE	χ^2	*df*	*p*
Intercept	-4.96	0.91			
Game (high risk)	-0.54	0.72	2.57	1	.11
Condition (no-barrier)	-0.10	0.82	0.69	1	.41
Order (no-barrier first)	-0.90	0.58	0.01	1	.91
Time waiting at stag	0.19	0.04	34.96	1	<.001
Trial number	-0.01	0.08	0.02	1	.90

Model 11: Checking back at hare. Model 11 analysed the likelihood of individuals to check back towards the stag or their partner, while they were still at hare. The overall test of the full model (including fixed factors: game (high or low risk), condition, order, sex, partner number and trial number; random effects: subject, pair, age; and random slopes: subjects across conditions, subjects across trials, pairs across conditions, and pairs across trials) compared to the reduced model (the same as the full model excluding test variables: game, condition, order and trial number) suggested that the test variables did significantly contribute to the likelihood to check back at hare (χ^2=41.70 , *df*=8, *p*<0.001, *N*=1536). The three way interaction between game, condition and order was not significant and was thus removed from the model, and in a second step the two-way interactions were also removed from the model. Table S6 is the results of the reduced model.

Table S6: The full results of the reduced model of checking back at hare.

Term (Test category)	Estimate	SE	χ^2	*df*	*p*
Intercept	-1.31	1.88			
Sex (male)	-1.58	1.01			
Partner number	-0.04	0.13			
Age	-0.03	0.04			
Game (high risk)	3.01	0.32	4.79	1	<.001
Condition (no-barrier)	0.02	0.42	0.00	1	.96
Order (no-barrier first)	0.61	0.35	3.06	1	.08
Trial number	-0.08	0.05	2.13	1	.14

Experiment 2b: High risk game with 4 year old children

Model 12: Decisions to leave hare. This model investigated the likelihood of an individual to leave hare on a given trial (6 decisions per individual). As with the chimpanzees, we expected there could be a difference in response to conditions across studies, and potentially the order in which they were experienced due to the modifications in the barrier condition in the high risk game; thus we included all interactions between game, condition and order in the full model. As we found no sex differences in Exp. 1, we did not test for these in these analyses, though we did include the term in the full and null models. The overall test of the full model (including fixed factors: game, condition, order, sex, trial number; random effects: individual and pair) compared to the reduced model (the same as the full model excluding test variables: game, condition, order, and the three way and all two way interactions between them; and trial number) suggested that the test variables did not significantly contribute to the likelihood to leave hare (χ^2=14.05, *df*=8, *p*=0.08, *N*=528).

Model 13: Coordination success. Model 13 investigated the factors involved in children's coordination success (how likely both individuals were to go for the stag on a given trial). We included all interactions between game, condition and order in the full model. The overall test of the full model (including fixed factors: game, condition, order, and trial number; random effects: pair; random slopes: pairs across conditions) compared to the reduced model (the same as the full model excluding test variables: game, condition, order, including all 3-way and 2-way interactions, and trial number) suggested that the test variables did not significantly contribute to the likelihood to coordinate successfully (χ^2=12.87, *df*=8, *p*=0.12, *N*=264).

Model 14: Communication at hare (individual decisions). Model 14 investigated the likelihood of children to communicate while are hare (6 trials per individual). Again, we included all interactions between game, condition and order in the full model. To see if the children changed their strategies across trials, we included trial number as a test variable. The

overall test of the full model of likelihood of individuals to communicate while still at hare (including fixed factors: game (high or low risk), condition, order, sex and trial number; random effects: subject, pair; and random slopes: subjects across trials) compared to the reduced model (full model excluding test variables: game, condition, order and trial number) suggested that the test variables did significantly contribute to the likelihood to communicate at hare (χ^2=31.28 , *df*=8, *p*<0.001, *N*=528). The three-way interaction between game, condition and order was not a significant contributor to the model and therefore removed, in a second step, all two-way interactions between these variables were also removed. Table S7 presents details of the reduced model including terms for the main effects of game, condition and order.

Table S7: The reduced model of communication at hare.

Term (Test category)	Estimate	SE	χ^2	*df*	*p*
Intercept	-2.93	0.67			
Sex (male)	-0.37	0.49			
Game (high risk)	2.40	0.53	24.70	1	<.001
Condition (no-barrier)	-0.19	0.32	0.36	1	.55
Order (no-barrier first)	0.01	0.48	0.00	1	.98
Trial number	-0.21	0.13	3.04	1	.08

Model 15: Communication while both at hare (pairs by trial). This model considered communication before either of the children in a pair had left their hare. For each trial we coded whether either of the children communicated verbally (6 trials per pair). As in the previous model we expected there could be a difference in response to conditions across studies, and the order in which they were experienced; thus we included all interactions

between game, condition and order in the full model. To see if the children changed their strategies across trials, we included trial number as a test variable. The overall test of the full model of likelihood of either child to communicate while both of them are still at hare (including fixed factors: game, condition, order, sex and trial number; random effects: pair; random slopes: pairs across trials) compared to the reduced model (full model: game, condition, order and trial number) suggested that the test variables did significantly contribute to the likelihood to communicate while both of them are still at hare (χ^2=63.36, *df*=8, *p*< .001, *N*=264). The three-way interaction between game, condition and order was not a significant contributor to the model and thus removed, in a second step, all two-way interactions between these variables were also removed. Table S8 is the reduced model including terms for the main effects of game, condition and order.

Table S8: The reduced model of communication at hare.

Term (Test category)	Estimate	SE	χ^2	*df*	*p*
Intercept	-3.70	0.94			
Sex (male)	-0.93	0.56			
Game (high risk)	4.36	0.83	60.50	1	< .001
Condition (no-barrier)	0.31	0.43	0.52	1	.47
Order (no-barrier first)	-0.25	0.60	0.17	1	.68
Trial number	-0.10	0.14	0.46	1	.50

Model 16: Communication at stag (individual decisions). Model 16 investigated the likelihood of children to communicate after leaving hare (6 trials per individual). Due to the modifications in the barrier condition in the high risk game we expected there could be a difference in response to conditions across studies, and potentially the order in which they were experienced; thus we included all interactions between game, condition and order in the

full model. To see if the children changed their strategies across trials, we included trial number as a test variable. The overall test of the full model of the likelihood of individuals to communicate after leaving hare (including fixed factors: game (high or low risk), condition, order, sex and trial number; random effects: subject, pair) compared to the reduced model (full model excluding test variables: game, condition, order and trial number) indicated that the test variables significantly contributed to the model (χ^2=21.58, *df*=8, p<0.01, *N*=488). Table S9 shows the results of the full model, with a significant three-way interaction between game, condition and order.

Table S9: The results of the full model of communication at stag.

Term (Test category)	**Estimate**	**SE**	χ^2	***df***	***p***
Intercept	-0.96	0.53			
Game (high risk)	1.41	0.56			
Condition (no-barrier)	0.12	0.61			
Order (no-barrier first)	1.31	0.70			
Sex (male)	0.13	0.34	0.18		.67
Trial number	-0.11	0.13	0.72		.40
Game*Condition	-1.28	0.62			
Game*Order	-2.35	0.81			
Condition*Order	-0.86	1.01			
Game *Condition *Order	2.29	0.88	6.74		.01

Model 17: Communication at stag (pairs by trial). This model examined only trials in which at least one child left their hare, and looked at the likelihood of pairs to communicate at stag (communication by either child; 6 trials per pair). Again, we expected there could be a

difference in response to conditions across studies, and potentially the order in which they were experienced; thus we included all interactions between game, condition and order in the full model and to see if they changed their strategies over time, we included trial number as a test variable. The overall test of the full model of the likelihood of individuals to communicate after leaving hare (including fixed factors: game (high or low risk), condition, order, sex and trial number; random effects: subject, pair) compared to the reduced model (full model excluding: game, condition, order and trial number) indicated that the test variables significantly contributed to the model (χ^2=23.39, df=8, p< .01, N=251). Table S10 shows the results of the full model, with a significant three-way interaction between game, condition and order.

Table S10: The results of the full model of communication at stag (pair measure).

Term (Test category)	Estimate	SE	χ^2	df	p
Intercept	0.82	0.68			
Game (high risk)	1.82	0.73			
Condition (no-barrier)	0.08	0.81			
Order (no-barrier first)	2.19	1.00			
Sex (male)	-0.26	0.41	0.39		0.53
Trial number	-0.10	0.19	0.29		0.59
Game*Condition	-1.74	0.86			
Game*Order	-4.07	1.14			
Condition*Order	-1.76	1.46			
Game *Condition *Order	3.96	1.28	10.12		.001

Model 18: Type of communication (content/attention-getters). This model examined the type of communication used by the children. We analysed only the trials in which individuals did communicate verbally and compared the likelihood to use content versus attention-getters. We included the three-way interaction between the phase (whether the children were still at hare or at stag), condition and game to investigate the potential differences in communication strategies across games in the model. Order of conditions was also included in the model, though not in the interaction, to reduce model complexity. The overall test of the full model of communication type: content/attention-getters (including fixed factors: game (high or low risk), phase (at hare or at stag), condition, order, sex and trial number; random effects: subject, pair; and random slopes: subjects across trials) compared to the reduced model (full model excluding test variables: game, phase, condition, and order) suggested that the proportion of content and attention-getters did not vary systematically across test variables (χ^2=12.90 , *df*=8, *p*=0.12, *N*=271).

Model 19:Type of communication (imperatives/informatives). With this model we aimed to look specifically at the type of content communication: whether children used imperatives or informatives conditionally in the different games, conditions and phases. We analysed only trials with content communication. To reduce model complexity we included only the two-way interactions between these factors, as well as a main effects for order and trial number. The overall test of the full model (including fixed factors: game (high or low risk), phase (at hare or at stag), condition, order, sex and trial number; random effects: subject, pair) compared to the reduced model (full model excluding test variables: game, phase, condition, order and trial number) indicated that the proportion of imperatives and informatives did vary systematically across test variables (χ^2=54.18, *df*=7, *p*<0.001, *N*=235). The two-way interactions did not significantly contribute to the model; Table S11 shows the reduced model with the main effects.

Table S11: The results of the full model of the type of content communication.

Term (Test category)	Estimate	SE	χ^2	*df*	*p*
Intercept	-0.51	0.83			
Sex (male)	0.14	0.53			
Game (high risk)	1.65	0.60	8.02	1	<.01
Phase (stag)	-2.27	0.49	29.96	1	<.001
Condition (no-barrier)	-0.86	0.39	5.26	1	.02
Order (no-barrier first)	-0.10	0.54	0.03	1	.86
Trial number	0.10	0.11	0.90	1	.34

Model 20: Monitoring of partner at hare. We investigated the likelihood of children to monitor their partner while still at hare (6 trials per individual). As it was not possible to monitor partner at hare in the barrier condition of the high risk game (due to the extended barrier) only no-barrier trials were analysed. We included the interaction term between game and order of conditions as well as including fixed factors: sex and trial number; random effects: subject, pair, and compared this to the reduced model (the same as the full model excluding test variables: game, order and trial number). This indicated that the test variables did contribute to the model (χ^2=10.81, *df*=3, *p*=0.01, *N*=264). The interaction between game and order did not significantly contribute to the model and was thus removed. Table S12 shows the results of the reduced model, suggesting that there was an increase in monitoring in the high risk game.

Table S12: The results of the full model of monitoring partner at hare.

Term (Test category)	**Estimate**	**SE**	χ^2	*df*	*p*
Intercept	0.16	1.13			
Sex (male)	0.22	0.47			
Game (high risk)	1.23	0.48	6.52	1	.01
Order (no-barrier first)	-1.33	0.79	2.78	1	.10
Trial number	-0.58	0.21	7.33	1	.01

Model 21: Monitoring partner at stag. We investigated the likelihood of children to monitor their partner after leaving hare (6 trials per individual). As there was decreased possibility to monitor partner in the barrier condition of the high risk game (due to the extended barrier, partner could only see each other when both were at stag) only no-barrier trials were analysed. The overall test of the full model (including fixed factors: game (high or low risk), order, sex and trial number; random effects: subject, pair) compared to the reduced model (the same as the full model excluding test variables: game, order and sex) suggested that the contribution of the test variables are marginal (χ^2=283.70, *df*= 3, *p*<0.001, *N*=243). The interaction between game and order did not significantly contribute to the model and was thus removed. Table S13 shows the results of the reduced model.

Table S13: The results of the full model of monitoring partner at stag.

Term (Test category)	**Estimate**	**SE**	χ^2	*df*	*p*
Intercept	1.40	1.00			
Game (high risk)	-1.85	0.38	21.69	1	<.001
Order (no-barrier first)	0.67	0.66	0.93	1	.33
Sex (male)	-0.99	0.38	6.35	1	.01
Trial number	0.01	0.18	0.00	1	.96

References

1. Baayen, R.H., D.J. Davidson, and D.M. Bates, *Mixed-effects modeling with crossed random effects for subjects and items.* Journal of Memory and Language, 2008. **59**(4): p. 390-412.
2. R Core Team, *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2012*. 2012, ISBN 3-900051-07-0.
3. Bates, D.M., *lme4: Mixed-effects modeling with R.* URL http://lme4. r-forge. r-project. org/book, 2010.
4. Forstmeier, W. and H. Schielzeth, *Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse.* Behavioral Ecology and Sociobiology, 2011. **65**(1): p. 47-55.